

|  |  |
| --- | --- |
| Recommendation systems  Integrated Continuous Assessment | Natalia de Oliveira Rodrigues Student ID 2023112  High Diploma in Science in Data analysis for Business |

**Table of Contents**

[Introduction 2](#_Toc152254415)

[The Purpose of Recommendation Systems in Machine Learning for Online Retail Businesses 3](#_Toc152254416)

[Comparative Analysis of Content and Collaborative Filtering: Machine Learning Models for Item-Item Collaborative Filtering 4](#_Toc152254417)

[Justification of Recommendations: Conceptual Insights for the Chosen Scenario 4](#_Toc152254418)

[Market Basket Analysis: Applying Apriori and FP-Growth Algorithms to Extract Insights 5](#_Toc152254419)

[Major Divergence Between Models: A Comparative Analysis of Machine Learning Results Using Apriori and FP-Growth Algorithms 6](#_Toc152254420)

[Designing an Interactive Dashboard for Older Adults (65+): Incorporating Features to Summarize Key Data Aspects and Justifying Dataset Suitability for Machine Learning Models in Online Retail 8](#_Toc152254421)

[Designing a User-Friendly Dashboard for Older Adults: Considerations and Strategies for Addressing Age-Related Challenges in Interface Design 10](#_Toc152254422)

[Conclusion 11](#_Toc152254423)

[Reference list 12](#_Toc152254424)

# Introduction

This academic paper aims to deliver a critical analysis of the knowledge produced, in the course: Higher Diploma in Science in Data Analytics for Business at CCT College. It integrates Machine Learning for Business, and Data Visualizations subjects.

This innovative project involves Content and Collaborative filtering techniques to develop a Recommendation System, and so provide personalized suggestions. Going beyond traditional recommendation methodologies, this paper also incorporates Market Basket Analysis to identify associations and enhance the recommendation process. For the Market Basket Analysis, we have applied Apriori and FP-Growth algorithms, and the result's comparison is provided in this paper.

To facilitate user interaction and understanding, we have worked in Python to develop a sophisticated and impressive interactive Uder-friendly dashboard that presents the system's insights and recommendations in an intuitive and visually appealing manner.

## The Purpose of Recommendation Systems in Machine Learning for Online Retail Businesses

According to (Wikipedia Contributors, 2019), online retail, also well known as eCommerce, allows customers to buy goods or services using web-based technology. Electronic Commerce permit businesses all over the world, the opportunity to collect customers' data. Technology makes it possible to build customers’ profiles based on demographic data, purchase history, preferences, and active data such as likes, views, clicks, and time spent.

According to the research (Bertens, Guitart and Chen Andáfrica Periáñez, 2018), E-commerce frequently uses recommendation systems, which are typically implemented using cooperative filtering techniques. Based on user ratings, they compare comparable products or users.

AI has helped businesses to build powerful analytics tools to improve customer satisfaction and drive revenue. Recommendation Systems enable businesses to learn about their customers and provide more meaningful content, increasing customer satisfaction, and driving business revenue (Nawrocka, Kot and Nawrocki, 2018).

Industry leaders in online retail are using recommendation systems. According to (Arkadiusz, 2021), Amazon’s recommendations are responsible for 35% of revenue, and Netflix’s recommendations are responsible for 80% of the movies seen on the platform. In 2023, according to (Cooper and McLachlan, 2023), 70% of the videos watched on YouTube were recommended to their customers.

Recommendation engines enable customers to find goods or services using different algorithms. They are Contant-based filtering, based on item features and user profile data; Collaborative-based filtering, based on the similarity with other users (user-to-user or item-to-item); and Hybrid systems that combine Contant-based filtering, and Collaborative-based filtering.

Figure 1 shows how content and collaborative techniques work in the context of a book recommendation system (Doshi, 2019).

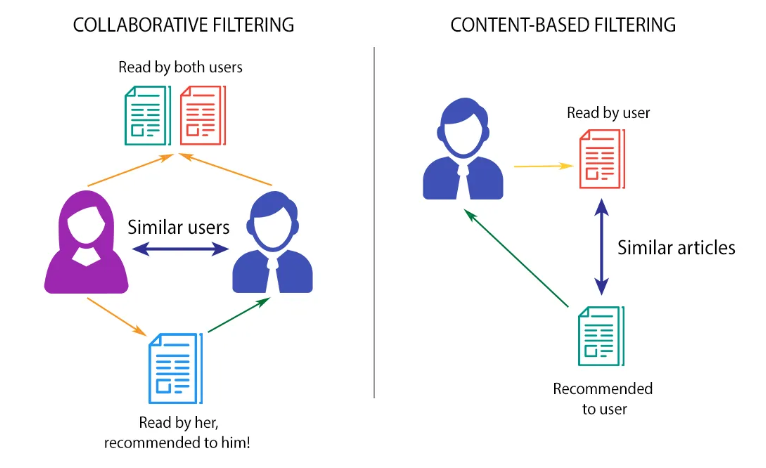


Figure Illustration of how Collaborative and Content Filtering works (Doshi, 2019)

## Comparative Analysis of Content and Collaborative Filtering: Machine Learning Models for Item-Item Collaborative Filtering

**Content Filtering Vs Collaborative Filtering:**

According to (Shivam, 2022), Content Filtering, recommends items based on the content similarity of items. It relies on the items’ characteristics. While Collaborative Filtering, recommends items based on user-item interactions, finding items that are similar in terms of user preferences. It is well known, that the choice between content-based and collaborative filtering depends on the specific use case and the available data.

For Content Filtering, TF-IDF (Term Frequency-Inverse Document Frequency) and sigmoid kernel for similarity computation were used as a method. In this paper, content filtering fails to recommend similar items based on the users’ preferences. All results from the recommendation list are like the given item. The chosen dataset does not bring information regarding items, only the names. The paper evaluates that the dataset is not suitable for content filtering, once it relies too much on item features.

For Collaborative Filtering, the cosine similarity method was used to build a user-item matrix between columns. Given an item 'whole milk', collaborative filtering recommends similar items based on the preferences of users who liked 'whole milk.' The similarity is calculated using the user-item interaction matrix. The results show the top 5 items that are most like 'whole milk' include 'other vegetables,' 'rolls/buns,' 'yogurt,' 'soda,' and 'tropical fruit.' It is successfully capturing personalized recommendations.

## Justification of Recommendations: Conceptual Insights for the Chosen Scenario

In conclusion, in the contant filtering results, it seems like the system is recommending the same item ('whole milk') rather than similar items. This is because the content-based filtering approach is designed to recommend items that are similar in content to the given item, and in this case, 'whole milk' is considered most like itself.

Given that, this paper recommends the use of Collaborative Filtering. It is effective in capturing personalized recommendations. Collaborative filtering is recommending items that are frequently co-purchased or interacted with by users who have also interacted with 'whole milk'. So, in collaborative filtering, you see a list of items that tend to be chosen by users who have chosen 'whole milk', and these items are different from 'whole milk' itself.

## Market Basket Analysis: Applying Apriori and FP-Growth Algorithms to Extract Insights

In retail organizations, market basket analysis is a crucial part of analytical CRM. Finding correlations or co-occurrences from transactional data through analysis can enable you to offer related products together, increasing revenue in the process. Many businesses have heavily relied on market basket analysis to identify product linkages and inform retailer promotion strategies (Charlet and Kumar, 2012).

According to (Manpreet and Shivani, 2016), affinity analysis or association rule learning are other names for market basket analysis (MBA). In the retail industry, it gives retailers the knowledge they need to comprehend customer purchasing patterns and make informed decisions.

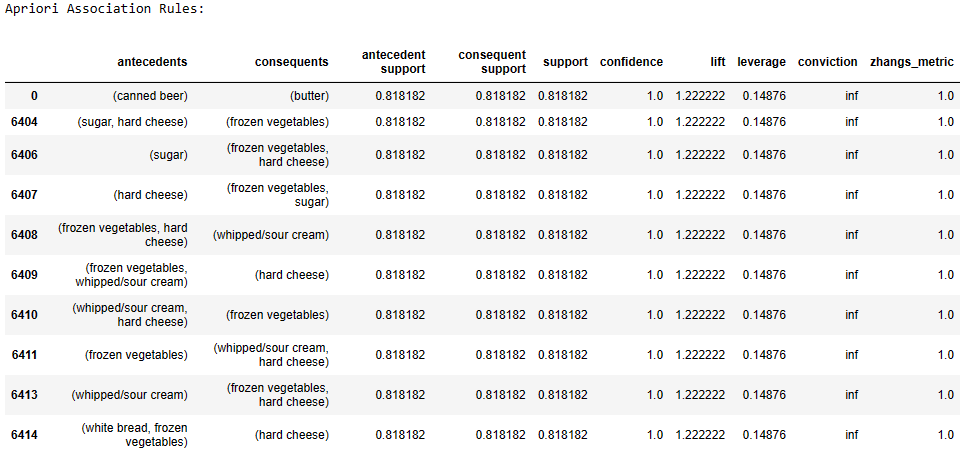


Figure Apriori Association Rules

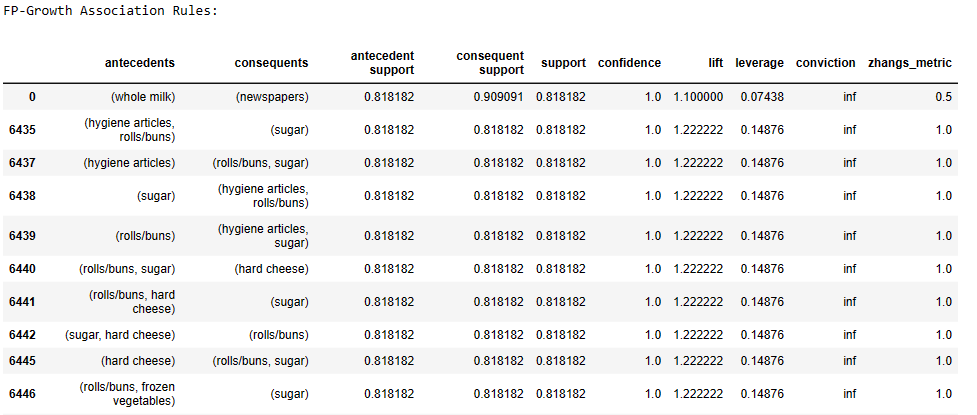


Figure FP-Growth Association Rules

## Major Divergence Between Models: A Comparative Analysis of Machine Learning Results Using Apriori and FP-Growth Algorithms

**The major divergence between models:**

Machine Learning Models Apriori and FP-Growth have the same main objective of identifying frequent itemsets in transaction datasets. Both approaches are to find patterns in the data, revealing important patterns in the data, in machine learning outputs. Apriori and FP-Growth are machine learning algorithms used for Market Basket analysis, but they have different approaches. Apriori is a kind of candidate generation method, that scans candidate itemsets until finds the frequent ones. In contrast, FP-Growth uses the frequency pattern on the data to growth approach, creating a tree, that is called FP-Tree. The tree represents the dataset and efficiently mining in two passes. This distinction leads to significant divergences in efficiency. This distinction leads to significant divergences in efficiency. When working with large datasets and sparse data, Apriori necessitates numerous database scans, which increases computational costs and memory use. In contrast, FP-Growth reduces runtime, memory requirements, and scan times, which makes it a better option for association rule mining jobs.

**Comparison and contrast the machine learning 10 first results:**

In this Market Basket Analysis, as given before, the association rules are created from two different algorithms, Apriori and FP-Growth. The results show the revealed associations between items in data based on their support, confidence, lift, and other metrics.

According to (Charlet and Kumar, 2012), frequent itemsets are used to generate association rules, with support and confidence serving as threshold values. The term "frequent itemset" refers to collections of items with the least amount of support. The percentage of transactions in the dataset that contain an itemset is known as the itemset’s support. The degree of assurance or dependability connected to each pattern that is found is known as confidence. The generated association rules rely on confidence.

In the Apriori results, the rules reveal strong associations between items such as "canned beer" and "butter," "sugar" and "hard cheese," and various combinations of "frozen vegetables," "whipped/sour cream," and "hard cheese."

On the other hand, the FP-Growth results show associations like "whole milk" and "newspapers," as well as patterns involving "hygiene articles," "rolls/buns," and "sugar." Like Apriori, the lift values are greater than 1, suggesting positive correlations.

For both algorithms, the lift values are consistently above 1, indicating positive associations. Lift is a metric that quantifies the degree to which the consequent is more likely than expected given the antecedent. A rule with a high lift indicates that buying the antecedent also greatly increases the likelihood of buying the consequent. A confidence of 0.8 means there is an 80% likelihood that the associated set of items. it measures the likelihood that the consequent item will be bought if the antecedent item is bought. The high confidence and lift indicate a strong association between the antecedent and the consequent.

In conclusion, it is crucial to note that, due to the algorithm’s methodologies, the specific associations and patterns discovered may vary. FP-Growth uses a frequent pattern growth strategy, while Apriori uses a candidate generation approach. FP-Growth shows to be more efficient, with quicker execution. Both algorithms reveal associations in the data, but the specific rules, efficiency of computation and memory usage are different from one to another.

## Designing an Interactive Dashboard for Older Adults (65+): Incorporating Features to Summarize Key Data Aspects and Justifying Dataset Suitability for Machine Learning Models in Online Retail

This project developed a sophisticated and impressive interactive Uder-friendly dashboard aimed at older adults (+65). The objective of the dashboard is to summarise important aspects of the data used to apply recommendation systems algorithms, insights, trends, and machine learning results.

Why old adults? In 2022, 771 million individuals worldwide were over 65 years of age or older, making up about 10% of the world's population (Alvarez, 2023). And according to (Tong, 2023) by the end of 2031, more than 25% of the workforce will be composed of older adults, in the G7 countries.

The Dashboard Header contains its title Groceries Sales Dashboard.

On the header left corner, below the title, it is possible to select the date range. It allows users to visualize any period wanted in the data.

On the central top section of the dashboard, users can utilize a dropdown to select their preferred frequency for data visualization including options like day, day of the week, and month.

On the header right corner, another dropdown allows users to pick a specific item. The resulting dash table, positioned in the top right grid, displays the top 10 items recommended by the Item-based Recommendation System analysed in the first part of this paper.

In addition to the recommendation table cited above, other three visualizations are provided: The Line graph showing the number of items sold that interacts with the date range picker and the data frequency dropdown. The Bar graph shows the Top 10 products sold, and the bottom 10 products sold table. Bar graph and bottom 10 table, both interact with the date range picker.

Users can get significant insights from the data in a quick navigation, such as:

* A noticeable increase in sales in 2015 compared to 2014.
* August exhibits the highest sales performance, while February represents the lowest performance in the year 2015.
* Wednesdays demonstrate the best sales performance, while Mondays indicate the lowest. This insight could drive decision-makers in scheduling advertising for Wednesdays when the supermarket is expected to be at full capacity and creating sales campaigns for Mondays to boost sales.
* The bar graph reveals the most sold items, with the ability to discern seasonal patterns if used in collaboration with the data range picker, such as items selling the most in December close to Christmas.
* The dash table showing the bottom 10 products sold assists decision-makers in evaluating products that may need to be discontinued or included in promotional deals to increase sales.
* The dash table generated by the item-based recommendation system facilitates quick assessment and comparison of recommended items.

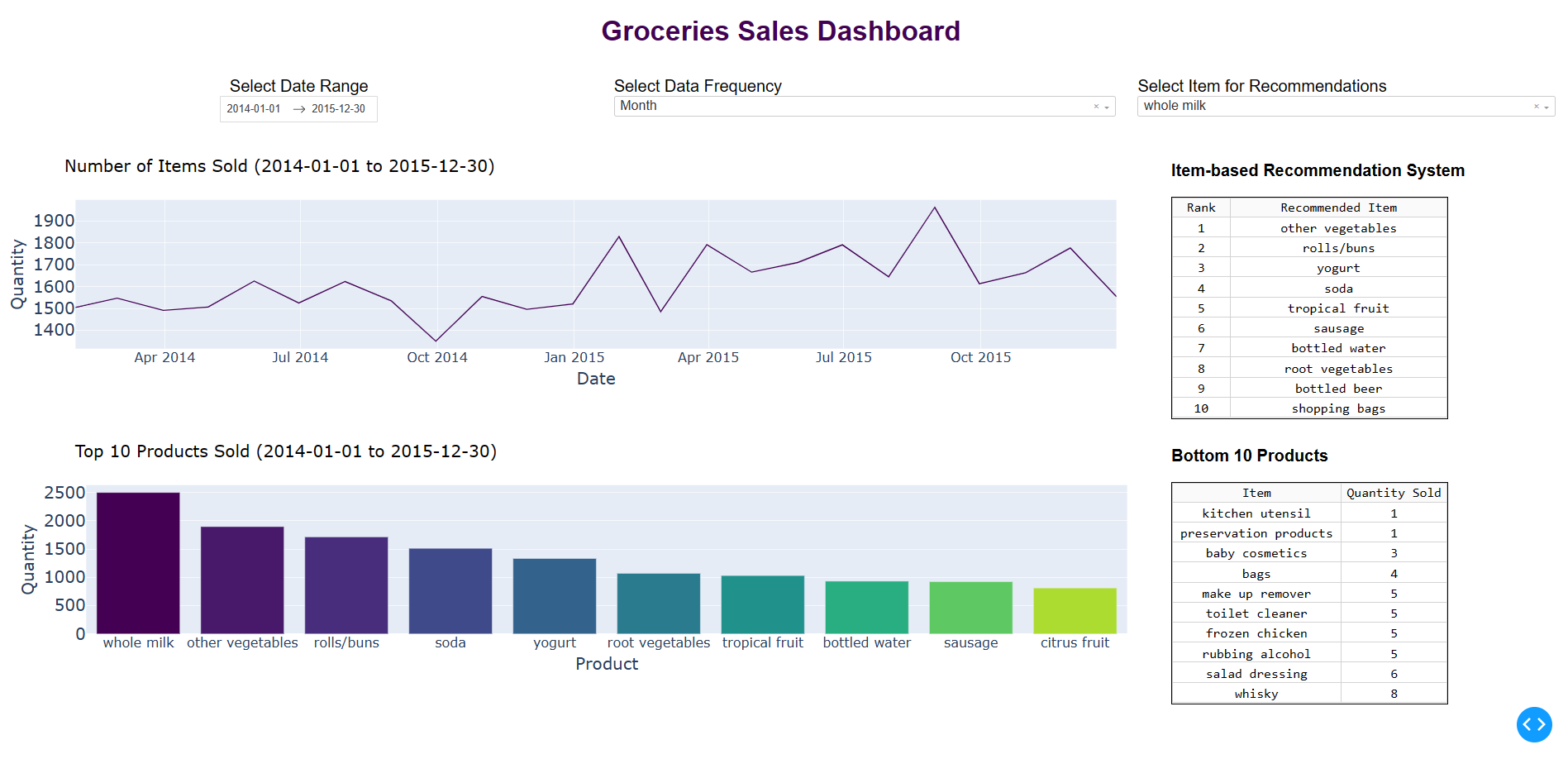


Figure Groceries Sales Dashboard

## Designing a User-Friendly Dashboard for Older Adults: Considerations and Strategies for Addressing Age-Related Challenges in Interface Design

It is well known, that ageing comes with inevitable physiological and cognitive changes. Design choices are made to overcome the challenges faced by older adults

According to (Polyuk, 2019) vision loss is what affect old user the most. It is important to consider visual accessibility when creating a visualization, especially increasing colour contrast. In this Dashboard project, the palette Viridis is used. This palette contributes to a positive user experience, offering smooth transitions between colours, and avoiding confusion in interpreting the data usually caused by abrupt changes.

(Polyuk, 2019) Also states that text and font size should be kept large as much as possible. Minimum of 16px. In this paper, large fonts are used. Especially for the dashboard title 50px and visualization titles 30px. Other font sizes are set as 24px.

To improve the interaction of older adults, this dashboard is sophisticated but simple. Containing gestures simple to perform such as simple horizontal, or vertical movements.

According to (Barros, Leitão and Ribeiro, 2014) It is recommended pickers, dropdowns, or checkboxes to the use of the keyboard be minimized. In this dashboard project, date ranger pick is added using dcc.DatePickerRange and two dropdowns are added as well, data frequency and item for recommendations dropdowns.

In short, the dashboard layout is easy to understand. It's neat, making it clear what's important in the data, and it gives valuable insights to help decision-makers.

# Conclusion

In conclusion, this academic paper explores the cross-disciplinary field of data analytics, combining Machine Learning and Data Visualization to produce a solid Recommendation System and Market Basket Analysis. This paper favours Collaborative filtering over Content filtering due its customized effectiveness. In the MBA making use of Apriori and FP-Growth algorithms, with FP-Growth proving more efficient for the chosen dataset.

In the second part, the paper introduces a stunning and sophisticated interactive dashboard, produced in Dash, adapted for older adults. The dashboard prioritizes visual accessibility with large fonts, a user-friendly interface, simplified gestures, and a clean layout. Addressing the challenges associated with aging, the Groceries Sales dashboard allow users to select data ranges, visualize wanted data frequencies, and display personalized recommendations. Providing valuable insights to help decision-makers promote informed strategies.

In heart, the project highlights the integration between advanced analytics, machine learning, and user-friendly interfaces in interpreting complex data related to the dynamic landscape of online retail.

# Reference list

Alvarez, P. (2023). *Charted: The World’s Aging Population from 1950 to 2100*. [online] Visual Capitalist. Available at: https://www.visualcapitalist.com/cp/charted-the-worlds-aging-population-1950-to-2100/ [Accessed 28 Nov. 2023].

Barros, A.C. de, Leitão, R. and Ribeiro, J. (2014). Design and Evaluation of a Mobile User Interface for Older Adults: Navigation, Interaction and Visual Design Recommendations. *Procedia Computer Science*, 27, pp.369–378. doi:https://doi.org/10.1016/j.procs.2014.02.041.

Ezz, M. and Elshenawy, A. (2019). Adaptive recommendation system using machine learning algorithms for predicting student’s best academic program. *Education and Information Technologies*, [online] 25(4), pp.2733–2746. doi:https://doi.org/10.1007/s10639-019-10049-7.

Khanal, S.S., Prasad, P.W.C., Alsadoon, A. and Maag, A. (2019). A systematic review: machine learning based recommendation systems for e-learning. *Education and Information Technologies*. [online] doi:https://doi.org/10.1007/s10639-019-10063-9.

Polyuk, S. (2019). *A guide to interface design for older adults*. [online] Toptal Design Blog. Available at: https://www.toptal.com/designers/ui/ui-design-for-older-adults [Accessed 25 Nov. 2023].

Portugal, I., Alencar, P. and Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, [online] 97, pp.205–227. doi:https://doi.org/10.1016/j.eswa.2017.12.020.

Roy, P.K., Chowdhary, S.S. and Bhatia, R. (2020). A Machine Learning approach for automation of Resume Recommendation system. *Procedia Computer Science*, [online] 167, pp.2318–2327. doi:https://doi.org/10.1016/j.procs.2020.03.284.

Tong, G.C. (2023). *‘Massive shift’: Older people will exceed a quarter of G7’s workforce by 2031, report says*. [online] CNBC. Available at: https://www.cnbc.com/2023/07/19/aging-workforce-older-people-to-fill-over-150-million-jobs-globally.html [Accessed 11 2023].